CEPSTRAL PEAK PROMINENCE-BASED PHONATION STABILISATION TIME AS AN INDICATOR OF VOICE DISORDER

Stephen Jannetts, Felix Schaeffler
Clinical Audiology, Speech and Language (CASL) Research Centre , Queen Margaret University, Edinburgh
EH12 8TS, Scotland, UK
SJannetts@qmu.ac.uk, FSchaeffler@qmu.ac.uk

I. INTRODUCTION

Impairment in the ability to initiate and sustain adequately periodic vocal fold vibrations is a common feature of voice disorders [3]. Traditional acoustic approaches that use sustained vowels in which initial/final portions are excluded have been criticised for poor validity and for exclusion of factors that may be a rich source of clinically relevant data [6, 4], for example regarding the onset of vocal fold vibration. Although rare, acoustic approaches to clinical analysis of voicing onset exist, e.g. vocal rise time [1]. However, these have not found widespread use in clinics or research.

The authors introduced a new measure of dynamic changes at onset of phonation during connected speech called phonation stabilisation time (PST) in a previous study [5]. Using autocorrelation values, the periodicity patterns at the onset of voiced segments of connected speech signals were analysed. PST was defined as the time taken for the autocorrelation values to rise from a voicing threshold to a stable threshold of periodicity. Results of this approach were promising but reliance on autocorrelation values might make the approach sensitive to pitch tracking artefacts. In the present study we therefore conducted a similar analysis, deriving PST from Cepstral Peak Prominence (CPP) rather than from the autocorrelation function.

The aim of this study was to establish if PST, as determined by CPP, is useful as an indicator of voice disorders in connected speech. We had three main hypotheses: that disordered voices would have greater (a) mean times and (b) standard deviation values than normal voices; and (c) that the proportion of voiced segments that reach the stable threshold of periodicity would be higher in normal voices.

To determine if PST is able to reveal deviations from periodicity in disordered voices that are not found by conventional acoustic analysis, we analysed a subset of voices for which there was a clinical diagnosis of voice disorder, but acoustic analysis of sustained vowels indicated that acoustic parameters were in the normal range. See [5] for details of selection of 'below threshold voices'.

II. METHODS

A. Material

The voice samples used in this study were obtained from the KayPENTAX Disordered Voice Database. Only samples from native English speakers, who had a valid diagnostic label and for whom both connected speech and sustained vowel data existed were selected. Table 1 shows sample demographics.

Table 1: Number of analysed sample by gender and health state

	Normal	Disordered	Total
Female	31	191	220
Male	21	121	142
Total	52	312	364

Connected speech samples consisted of the first 12 seconds of 'The Rainbow Passage'. All sound files were resampled to 10 kHz prior to further analysis.

B. Acoustic analysis

All sound files were segmented into voiced/unvoiced portions using Praat (version 5.4.08). CPP values were derived using VoiceSauce. Only voiced segments of 70 ms or longer were considered for further analysis [2]. For these segments, the Phonation Stabilisation Time (PST) was calculated as the duration between the beginning of the voiced segment and a threshold value. A threshold value of 23.14 dB was determined in a pilot study using manual segmentations of signals from 10 normal speakers, this value being the mean CPP for all normal speakers.

Any segments for which the stable periodicity threshold was not reached were not included in PST calculations. However, the proportion of voiced segments reaching threshold for every speaker was recorded as a percentage (named 'Seg%' in the following).

C. Statistical analysis

The two gender groups were analysed separately. Mann-Whitney U-tests were used to compare mean ranks.

III. Results

Summary statistics for each of the three variables investigated are shown in Table 2. Mann-Whitney U-test results are presented in Table 3. Note that the 'below threshold' groups are subsets of the main groups, as described above. Disordered voices from all groups showed a significantly longer mean duration than normal voices from the same group. Similarly, SD of PST was significantly larger for disordered voices in all groups. The proportion of voiced segments that reached the stable threshold of periodicity were significantly higher for normal voices in all groups. Fig. 1 illustrates the distribution of mean PST values for all groups.

Table 2: Descriptive statistics (Mean(SD)) for each variable for all groups.

	Female Group		Male Group		Female Group (Below Threshold)		Male Group (Below Threshold)	
	Normal	Disordered	Normal	Disordered	Normal (n = 30)	Disordered (n = 17)	Normal (n = 14)	Disordered (n = 16)
PST M	51.1(11.3)	82.9(39.2)	45.9(9.5)	77.89(33)	50.8(11.4)	70.3(27.2)	46.6(9.3)	93.1(47)
PST SD	23.6(7.8)	58.9(39.2)	26.9(10)	49.3(25.2)	23.5(7.9)	52.1(28.5)	26.1(9.6)	65.4(37.8)
Seg%	93.6(6.2)	55.8(33.2)	95.6(4.4)	76.6(22.1)	93.6(6.3)	79(23.7)	95.4(4.9)	52.9(32.8)

Table 3: Mann-Whitney U values for normal vs. disordered voices for each variable for all groups.

	Female Group	Male Group	Female Group (Below Threshold)	Male Group (Below Threshold)
PST M	728.5**	221.5**	105.5*	41*
PST SD	502**	200**	64**	42*
Seg%	557.5**	140**	145.5*	21.5**

*p<0.005, **p<0.001

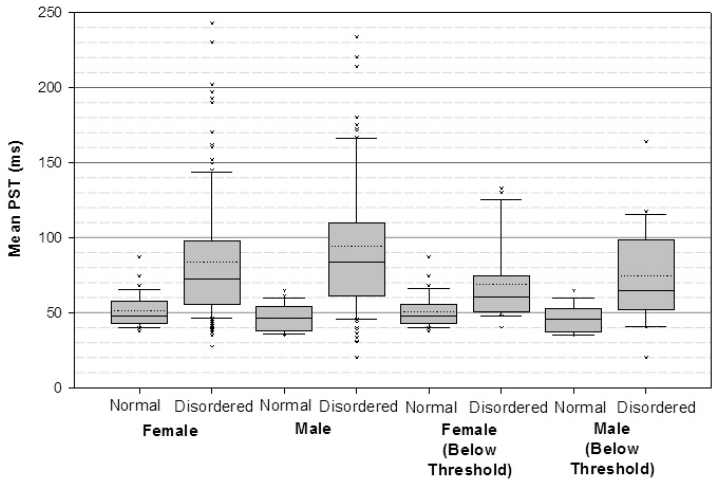

Fig. 1: Boxplots of mean PST for all groups

IV. Discussion

PST using autocorrelation has already shown potential in differentiating between normal and disordered voices [5]. The aim of this study was to establish if CPP-based PST is also useful as an indicator of voice disorders in connected speech, and might outperform PST based on autocorrelation. Our results indicate that PST using CPP has potential to differentiate between the normal and disordered voices from the Voice Disorder Database, and outperforms our previous measure, especially regarding male voices. The results for the 'below threshold' groups for both male and female are of particular interest. These results suggest that PST using CPP may be a potential indicator of voice disorder in cases where traditional acoustic analysis of sustained vowels do not show any pathological findings.

[1] Baken, R.J. & Orlikoff, R.F., 2000. *Clinical measurement of speech and voice*, San Diego: Singular Publishing.

[2] Crystal, T.H. & House, A.S., 1988. Segmental durations in connected-speech signals: Current results. *The journal of the acoustical society of America*, 83(4), pp.1553–1573.

[3] Gordon, M. & Ladefoged, P., 2001. Phonation types: a cross-linguistic overview. *Journal of Phonetics*, 29(4), pp.383–406.

[4] Maryn, Y. & Roy, N., 2012. Sustained vowels and continuous speech in the auditory-perceptual evaluation of dysphonia severity. *Jornal da Sociedade Brasileira de Fonoaudiologia*, 24(2), pp.107–12.

[5] Schaeffler, F., Beck, J. & Jannetts, S., 2015. Phonation Stabilisation Time as an Indicator of Voice Disorder. *ICPhS [submitted]*.

[6] Takahashi, H. & Koike, Y., 1976. Some perceptual dimensions and acoustical correlates of pathologic voices. *Acta oto-laryngologica. Supplementum*, 338, pp.1–24.